

Does sequential augmenting of the simple linear heteroscedastic regression reduce variances of the Ordinary Least Squares?*

Andrzej S. Kozek^{*a†} and Brian Jersky^b

Macquarie University^a, Sydney and California State Polytechnic University^b, Pomona

2012-03-22

Abstract

If uncorrelated random variables have a common expected value and decreasing variances then the variance of a sample mean is decreasing with the number of observations. Unfortunately, this natural and desirable Variance Reduction Property (VRP) by augmenting data is not automatically inherited by Ordinary Least Squares (OLS) estimators of parameters. In the paper we find conditions for the OLS to have the VRP. In the case of a straight line regression we show that the OLS estimators of intercept and slope have the VRP if the design points are increasing. This also holds true for alternating two-point experimental designs. The obtained results are useful in the cases

*The final and revised version of the paper has been accepted for publication in *Statistics*, ISSN: 0233-

where it is known that variances of the subsequent observations are non-increasing, but the ratios of the decrease are not available to use sub-optimal or optimal Weighted Least Squares estimators.

AMS (2010) subject classification. Primary 62J05; secondary 62K99, 15A45, 62G20.

Keywords and phrases. Augmented data, design of experiment, linear models, ordinary least squares, regression, variance reduction.

1 Introduction

If uncorrelated random variables have a common expected value and decreasing variances then, in agreement with common sense, the variance of a sample mean is decreasing with the number of observations. Surprisingly, this natural and desirable Variance Reduction Property (VRP) resulting from inclusion of less contaminated data is not shared in general by the Ordinary Least Squares (OLS) estimators of parameters.

In fact, by generating random regression designs, it is easy to come across experimental designs where by adding additional observation with a smaller error variance, we in fact increase the variance of the augmented OLS estimators (we refer to Example 3 for a specific case). On the other hand, many such random designs indeed enjoy the VRP. In the case of homoscedastic models with identical variances, or when the covariance matrix is known up to a constant, the VRP is well known, with the Weighted Least Squares replacing the OLS in the latter case, and a fairly complete theory on optimal augmenting of the data is then available. We refer the reader to Sengupta (1995) and Bhaumik & Mathew (2001) for the most comprehensive recent results in this direction, and for further references to the related

literature. The case of unknown and decreasing variances appears difficult and seemingly has eluded the attention of researchers. This case however covers important applications in the theory of Asymptotic Statistics, where the asymptotic distributions of smoothed functionals are exactly of such a form.

In the paper we derive (Theorem 2) general conditions characterizing designs for which the OLS estimators of parameters retain the variance reduction property when new data with lower variances are included. These criteria can be easily used to iteratively construct sequences of experimental designs with the VRP. We found one interesting general case where the VRP remains valid: when n uncorrelated observations have equal variances of errors and the $(n + 1)$ -th observation has a lower variance. Then the variance of the OLS estimator of parameters based on $(n + 1)$ observations is lower than that based on n observations (Theorem 1). Following the method of Theorem 2 in a particularly important linear model with two parameters: the intercept and slope, we show then that any design with the explanatory variable increasing with n has the VRP property (Theorem 3), and that any two-point alternating experimental design has the VRP property (Theorem 4). All these results are non-trivial, novel and guarantee non-negativeness of the diagonal elements of matrices which themselves need not be positively definite.

The paper is organized in the following way. In Section 2 we present the main results of the paper. Section 3 contains details of the straight line regression model and proofs of Theorems 3 and 4. In the proofs we include only the main nontrivial steps and omit the detailed, often long, though elementary algebraic manipulations. Technical details can be found in Appendix A. In Section 4 we include examples illustrating that the assumptions of Theorem 3 cannot be weakened. In Appendix A technical auxiliary results needed in

the proofs of the main theorems of the paper are presented. We also include Proposition 4 showing decomposition of variances in the general case of dependent random variables. Proposition 4 extends the decomposition (14) valid for uncorrelated random variables. We acknowledge using SCIENTIFIC WORKPLACE for deriving the results of this paper.

2 Variance reduction in heteroscedastic models

Consider the linear model

$$\mathbf{Y}_0 = A_0\beta + \epsilon_0, \quad (1)$$

where A_0 is an (n, k) design matrix, β is a $(k, 1)$ column vector of regression parameters, and $\mathbf{Y}_0 = (Y_1, Y_2, \dots, Y_n)^T$ and $\epsilon_0 \in R^n$ are random vectors. We also consider an augmented model

$$\mathbf{Y}_1 = A_1\beta + \epsilon_1, \quad (2)$$

where $A_1 = \begin{bmatrix} A_0 \\ a \end{bmatrix}$ is an $(n+1, k)$ design matrix with a being a design of the $(n+1)$ -th experiment, $\mathbf{Y}_1 = [\mathbf{Y}_0^T, Y_{n+1}]^T$ and $\epsilon = [\epsilon_0^T, \epsilon_{n+1}]^T \in R^{n+1}$ are random vectors such that

$$E(\epsilon_1) = 0, \quad Cov(\epsilon_1) = \Sigma_{11}. \quad (3)$$

Apart from Example 2 and Proposition 4 we will consider uncorrelated random variables with a diagonal covariance matrix Σ_{11} . Hence we have

$$\Sigma_{11} = \begin{bmatrix} \Sigma_{00} & O \\ O^T & \sigma_{n+1}^2 \end{bmatrix}, \quad (4)$$

where O is an $(n, 1)$ vector of zeros and Σ_{00} is a diagonal covariance matrix of \mathbf{Y}_0 . To simplify notation we define

$$A_{11} = A_1^T A_1 \quad \text{and} \quad (5)$$

$$A_{00} = A_0^T A_0, \quad (6)$$

respectively. Both β and the covariance matrix Σ_{11} are assumed to be unknown and the design matrix A_1 is assumed to be of full rank. The OLS estimators of parameters β based on models (1) and (2) are of the form

$$\hat{\beta}_1 = A_{11}^{-1} A_1^T \mathbf{Y}_1 \quad (7)$$

and

$$\hat{\beta}_0 = A_{00}^{-1} A_0^T \mathbf{Y}_0, \quad (8)$$

and have covariance matrices V_{11} and V_{00} given by

$$V_{11} = A_{11}^{-1} A_1^T \Sigma_{11} A_1 A_{11}^{-1} \quad (9)$$

and

$$V_{00} = A_{00}^{-1} A_0^T \Sigma_{00} A_0 A_{00}^{-1}, \quad (10)$$

respectively. In the paper we derive conditions implying inequalities

$$V_{11}(i, i) \leq V_{00}(i, i) \quad \text{for } i \in \{1, 2, \dots, k\}, \quad (11)$$

where V_{11} and V_{00} are given by (9) and (10), respectively. We will refer to property (11) as the variance reduction by augmenting the data, or, for short, as the VRP property.

It will be convenient to use the following notation throughout the paper

$$d_{n+1} = aA_{00}^{-1}a^T, \quad \text{and} \quad (12)$$

$$q_{n+1} = 1 + d_{n+1}. \quad (13)$$

In the general case of a heteroscedastic model we have the following decomposition of variance V_{11} .

Proposition 1 *If the covariance matrix Σ_{11} and design matrix A_1 in the model given by (2) are both of full rank, then for uncorrelated random variables Y_i , $i = 1, \dots, n+1$, we have*

$$V_{11} = V_{00} - \sigma_{n+1}^2 W - W_{11}, \quad (14)$$

where

$$W = \frac{1}{q_{n+1}} A_{00}^{-1} a^T a A_{00}^{-1}, \quad (15)$$

$$W_{11} = A_{00}^{-1} D_{11} A_{00}^{-1} - A_{11}^{-1} D_{11} A_{11}^{-1}, \quad (16)$$

$$D_{11} = A_0^T D A_0, \quad \text{and} \quad (17)$$

$$D = \Sigma_{00} - \sigma_{n+1}^2 I. \quad (18)$$

In the particular case of iid random variables with common variance σ^2 the matrices D and W_{11} equal zero and then (14) reduces to the well known case of the Plackett (1950) updating formula for covariances of the LS estimators

$$A_{11}^{-1} = A_{00}^{-1} - W. \quad (19)$$

Let us note that in (14) the matrix $\sigma_{n+1}^2 W$ is non-negative definite and hence it always reduces the variances on the diagonal of V_{11} . However, in general, the matrix W_{11} need not be

positive definite, cf. Example 1 in Section 4, and for general designs it may even have negative elements on the diagonal. Consequently, in the heteroscedastic case the variance of estimators of β can increase with an increase of the number of observations n , even when k and the parameters β are kept the same. As might be expected, if $\sigma_{n+1} > \max\{\sigma_i, i = 1, \dots, n\}$, then the variance of $\hat{\beta}_1$ could be greater than that of $\hat{\beta}_0$. Therefore in the following we shall concentrate on the cases where the covariance matrices

$$\Sigma_{11} = \text{Diag}\{\sigma_1^2, \sigma_2^2, \dots, \sigma_{n+1}^2\} \quad (20)$$

have non-increasing standard deviations $\sigma_i, i = 1, 2, \dots, n+1$. Clearly, in these cases the matrix D given by (18) is also diagonal with non-increasing diagonal elements $D_i = \sigma_i^2 - \sigma_{n+1}^2 \geq 0$. As has been mentioned previously in the Introduction, decreasing σ_i 's alone do not guarantee the VRP, however.

Nonetheless, we have the following surprising general case where the VRP is valid: if the first n observations have uncorrelated errors with the same variances and the error in the $(n+1)$ -th uncorrelated experiment is lower than in the preceding cases, then the ordinary least squares estimators based on the augmented data have lower variances than in the case of n observations and the homoscedastic model.

Theorem 1 *If $D = \sigma^2 I$ where I is the identity matrix then*

$$W_{11} = \sigma^2 \left(2 - \frac{d_{n+1}}{1 + d_{n+1}} \right) W \quad (21)$$

and, consequently, W_{11} is non-negative definite. \diamond

Clearly Theorem 1 and Proposition 1 imply the following corollary.

Corollary 1 *If random variables Y_1, Y_2, \dots, Y_{n+1} are uncorrelated, $\text{var}(Y_i) = \sigma_1^2$ for $i = 1, 2, \dots, n$ and $\text{var}(Y_{n+1}) = \sigma_{n+1}^2$ with $\sigma_1^2 > \sigma_{n+1}^2$ then the OLS estimators have the variance reduction property. \diamond*

The proof of Theorem 1 can be obtained using (57) and (58) and some straightforward algebra, and is omitted.

In the general uncorrelated case we have the following characterization of the non-negative diagonal elements of W_{11} , implying the VRP.

Proposition 2 *The diagonal element $W_{11}(i, i)$ of W_{11} is non-negative for every diagonal covariance matrix (20) with σ_i non-increasing if and only if*

$$\sum_{j=1}^m \left(\left(\sum_{l=1}^k A_{00}^{-1}(i, l) A_0(j, l) \right)^2 - \left(\sum_{l=1}^k A_{11}^{-1}(i, l) A_0(j, l) \right)^2 \right) \geq 0 \quad \text{for } m = 1, 2, \dots, n. \quad (22)$$

The following theorem follows immediately from Propositions 1 and 2.

Theorem 2 *The OLS estimator of the i -th component of β has the VRP if condition (22) is met.*

Conditions (22) are in general not easy to verify. However, given the design matrix A_1 of the first n experiments one can use (19) to find numerically designs a for the $(n+1)$ -st experiment satisfying inequalities (22). Hence Proposition 2 provides a method for a sequential numerical construction of experimental designs having the VRP. Remark 3 in Section 3 provides more details in the case of a straight line linear model.

The linear model for a straight line regression has the following design matrix for $n > 1$.

$$A_1 = \begin{bmatrix} \mathbf{1} & \mathbf{h}_{n+1} \end{bmatrix} = \begin{bmatrix} 1 & h_1 \\ \vdots & \vdots \\ 1 & h_n \\ 1 & h_{n+1} \end{bmatrix}. \quad (23)$$

The following two theorems show that in this important case the VRP remains a valid feature of two important classes of regression designs: with non-negative and increasing explanatory variables h_i (Theorem 3) and with two alternating values of the explanatory variable (Theorem 4).

Theorem 3 *If random variables $Y_i, i = 1, 2, \dots, n+1$ are uncorrelated with non-increasing variances σ_i^2 and with increasing values of the explanatory variable $h_i \geq 0$ then the OLS estimators with design matrix (23) have the variance reduction property (11).*

Remark 1 *In Theorem 3 we consider models with a simple explanatory variable h . Clearly, any increasing function of h can be used here, as such a model, by applying a suitable change of parametrization, can be easily transformed to the case with a design matrix (23).*

Examples 1 – 3 of Section 4 show that the assumptions of Theorem 3 cannot be weakened. This however does not contradict the fact that the VRP may also hold for many regression designs which do not meet the monotonicity assumption of Theorem 3. For example, it is well known that in the homoscedastic case two-point simple regression designs result in the lowest variance of the LS estimators of the slope and intercept. The following theorem shows that these two-point regression designs also have the variance reduction property.

Theorem 4 *If random variables $Y_i, i = 1, 2, \dots, n + 1$ are uncorrelated with variances σ_i^2 non-increasing with i , with two distinct values of the explanatory variable given by*

$$h_i = \begin{cases} b & \text{if } i \text{ is odd} \\ c & \text{if } i \text{ is even} \end{cases}$$

then the OLS estimators (8) with design matrix (23) and non-increasing variances (20) have the variance reduction property (11).

3 Variance reduction for a straight line model

Consider the case of the straight line model (23). By Theorem 2 to prove the variance reduction property we need to verify that conditions (22) hold true. In this particular case (22) is equivalent to inequalities (43) and (43) of Condition C3. In Theorem 5 we show implication relations between Conditions C1 — C7 given in Subsection 3.2. In the proofs of Theorems 3 and 4 we show that these conditions are met under the conditions specified in the corresponding theorems. We need however first to work out the structure of the diagonal elements of the matrix W_{11} .

3.1 The Diagonal of Matrix W_{11}

It will be convenient to use the following notation

$$S_{1,n} = \sum_{i=1}^n h_i, \quad S_{2,n} = \sum_{i=1}^n h_i^2, \quad \text{and} \tag{24}$$

$$V_n = \frac{S_{2,n}}{n} - \left(\frac{S_{1,n}}{n} \right)^2 = \frac{1}{n} \sum_{i=1}^n \left(h_i - \frac{S_{1,n}}{n} \right)^2. \tag{25}$$

Notice that if $S_{1,n} \neq 0$ then (25) is equivalent to

$$\frac{S_{2,n}}{S_{1,n}} - \frac{S_{1,n}}{n} = \frac{V_n}{S_{1,n}/n}. \quad (26)$$

Moreover, let us note that d_{n+1} given by (12) reduces now to

$$d_{n+1} = \frac{1}{n} \left(\frac{(S_{1,n} - nh_{n+1})^2}{n^2 V_n} + 1 \right). \quad (27)$$

Matrices A_{00}^{-1} and W reduce in the present case to the following form

$$A_{00}^{-1} = \frac{1}{n^2 V_n} \begin{bmatrix} S_{2,n} & -S_{1,n} \\ -S_{1,n} & n \end{bmatrix}, \quad (28)$$

$$W = \gamma_W \begin{bmatrix} (S_{2,n} - S_{1,n}h_{n+1})^2 & -(S_{2,n} - S_{1,n}h_{n+1})(S_{1,n} - nh_{n+1}) \\ -(S_{2,n} - S_{1,n}h_{n+1})(S_{1,n} - nh_{n+1}) & (S_{1,n} - nh_{n+1})^2 \end{bmatrix}, \quad (29)$$

where

$$\gamma_W = \frac{1}{q_{n+1}} \left(\frac{1}{n^2 V_n} \right)^2. \quad (30)$$

and q_{n+1} is given by (13). We also get

$$A_{11}^{-1} = A_{00}^{-1} - W = \frac{1}{q_{n+1}} \frac{1}{n^2 V_n} \begin{bmatrix} S_{2,n+1} & -S_{1,n+1} \\ -S_{1,n+1} & n+1 \end{bmatrix} \quad (31)$$

and

$$A_0^T D A_0 = D_{11} = \begin{bmatrix} \delta_{11} & \delta_{1h} \\ \delta_{1h} & \delta_{hh} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n D_i & \sum_{i=1}^n D_i h_i \\ \sum_{i=1}^n D_i h_i & \sum_{i=1}^n D_i h_i^2 \end{bmatrix}, \quad (32)$$

where D_{11} is given by (17) and D_i are the diagonal elements of matrix D given by (18). By

applying (19) we infer from (28), (29) and (31) that q_{n+1} given by (13) equals

$$q_{n+1} = \frac{(n+1)^2 V_{n+1}}{n^2 V_n}, \quad (33)$$

where V_{n+1} stands for the *variance* of the design points given by (25) with substitution $n \leftarrow n + 1$. Clearly, since $D_i \geq 0$ the matrix $A_0^T D A_0$ is non-negative definite.

Remark 2 *It may be interesting to note that (33) combined with (13) and (27) is equivalent to the Welford (1962) updating formula for $S_{2,n}$.*

It will be convenient to use the following notation.

$$\begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \end{pmatrix} = \begin{pmatrix} q_{n+1}S_{2,n} \\ S_{2,n+1} \\ q_{n+1}S_{1,n} \\ S_{1,n+1} \\ nq_{n+1} \\ n+1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \alpha_4 \\ \alpha_5 \end{pmatrix} = \begin{pmatrix} a_1^2 - a_2^2 \\ a_1a_3 - a_2a_4 \\ a_3^2 - a_4^2 \\ a_3a_5 - a_4a_6 \\ a_5^2 - a_6^2 \end{pmatrix}. \quad (34)$$

We hasten to note that in the present case conditions (22) are equivalent to (45) -(46), respectively. However, to get conditions equivalent to (22) in a simpler algebraic form we begin with the following diagonal form of W_{ii} .

Proposition 3 *The diagonal elements of W_{11} are given by*

$$W_{11}(1, 1) = \frac{\alpha_1\delta_{11} - 2\alpha_2\delta_{1h} + \alpha_3\delta_{hh}}{(n+1)^4V_{n+1}^2} \quad (35)$$

and

$$W_{11}(2, 2) = \frac{\alpha_3\delta_{11} - 2\alpha_4\delta_{1h} + \alpha_5\delta_{hh}}{(n+1)^4V_{n+1}^2}, \quad (36)$$

respectively, where δ_{ij} are given by (32). \diamond

Prior to further considering conditions guaranteeing non-negativeness of the diagonals of W_{11} we shall need some properties of polynomials driving the behavior of numerators in (35) and (36).

3.2 Driving polynomials

Let us introduce two closely related pairs of quadratic polynomials: $(p_1(h), p_2(h))$ and $(\bar{p}_1(h), \bar{p}_2(h))$, referred to in the following as driving polynomials. We need them to formulate conditions equivalent to the variance reduction property. Polynomials $p_1(h)$ and $p_2(h)$ are given by

$$p_1(h) = \alpha_1 - 2\alpha_2 h + \alpha_3 h^2 \quad \text{and} \quad (37)$$

$$p_2(h) = \alpha_3 - 2\alpha_4 h + \alpha_5 h^2, \quad (38)$$

respectively, where $\alpha_1, \alpha_2, \dots, \alpha_5$ are given by (34). In Lemmas 2 and 3 we derive their roots $r_{1,1}, r_{1,2}$ and $r_{2,1}, r_{2,2}$ given by (64)–(67), respectively and show features needed in the sequel. Polynomials $\bar{p}_1(h)$ and $\bar{p}_2(h)$ are given by

$$\bar{p}_1(h) = (h - r_{1,1})(h - r_{1,2}) \quad \text{and} \quad (39)$$

$$\bar{p}_2(h) = (h - r_{2,1})(h - r_{2,2}) \quad (40)$$

and are obtained from $p_1(h)$ and $p_2(h)$ by dividing them by α_3 and α_5 , respectively.

3.3 Necessary and sufficient conditions for the VRP.

In the following we consider statements **C1–C7** and the corresponding assumptions under which they are equivalent to the non-negativeness of the diagonal of W_{11} .

C1 The diagonal elements of matrix W_{11} given by (35)-(36) are non-negative for any vector of non-increasing and non-negative D_i 's.

C2 For all non-increasing and non-negative D_i 's

$$\alpha_1 \delta_{11} - 2\alpha_2 \delta_{1h} + \alpha_3 \delta_{hh} \geq 0, \quad \text{and} \quad (41)$$

$$\alpha_3 \delta_{11} - 2\alpha_4 \delta_{1h} + \alpha_5 \delta_{hh} \geq 0. \quad (42)$$

C3 For all non-increasing and non-negative D_i 's

$$\sum_{i=1}^n D_i (\alpha_1 - 2\alpha_2 h_i + \alpha_3 h_i^2) \geq 0 \quad \text{and} \quad (43)$$

$$\sum_{i=1}^n D_i (\alpha_3 - 2\alpha_4 h_i + \alpha_5 h_i^2) \geq 0. \quad (44)$$

C4 For each $m \in \{1, \dots, n\}$

$$\alpha_1 - 2\alpha_2 \frac{S_{1,m}}{m} + \alpha_3 \frac{S_{2,m}}{m} \geq 0, \quad \text{and} \quad (45)$$

$$\alpha_3 - 2\alpha_4 \frac{S_{1,m}}{m} + \alpha_5 \frac{S_{2,m}}{m} \geq 0. \quad (46)$$

C5 For each $m \in \{1, \dots, n\}$ we have

$$p_1 \left(\frac{S_{1,m}}{m} \right) + \alpha_3 V_m \geq 0, \quad \text{and} \quad (47)$$

$$p_2 \left(\frac{S_{1,m}}{m} \right) + \alpha_5 V_m \geq 0, \quad (48)$$

where polynomials $p_1(h)$ and $p_2(h)$ are given by (37) and (38), respectively.

C6 For each $m \in \{1, \dots, n\}$ we have

$$\bar{p}_1 \left(\frac{S_{1,m}}{m} \right) + V_m \geq 0, \quad \text{and} \quad (49)$$

$$\bar{p}_2 \left(\frac{S_{1,m}}{m} \right) + V_m \geq 0, \quad (50)$$

where polynomials $\bar{p}_1(h)$ and $\bar{p}_2(h)$ are given by (39) and (40), respectively.

C7 For each $m \in \{1, \dots, n\}$

$$\bar{p}_1 \left(\frac{S_{1,m}}{m} \right) + V_m \geq 0. \quad (51)$$

Theorem 5 1. If $V_n > 0$ then statements **C1** – **C5** imply each other.

2. If V_n, α_3 and α_5 are positive then **C1** – **C6** imply each other.

3. If h_1, \dots, h_{n+1} are non-negative and increasing then **C1** – **C7** imply each other.

Proof of Theorem 5.

Assuming that $V_n > 0$, the equivalence of **C1**, **C2** and **C3** is evident. Lemma 1 applied with $U_i = \alpha_1 - 2\alpha_2 h_i + \alpha_3 h_i^2$ implies the equivalence of **C3** and **C4**. By applying (25) we get equivalence of **C4** and **C5**. Assuming additionally that $\alpha_3 > 0$ and $\alpha_5 > 0$ we can divide (47) by α_3 and (48) by α_5 , respectively, without changing the direction of the inequalities. This shows equivalence of **C5** and **C6**. If h_1, \dots, h_{n+1} are non-negative and increasing then by (68)–(70), Lemma 7 and Corollaries 2 and 3 both polynomials have the same roots $r_{1,1} = r_{2,1}$. Moreover, the root $r_{1,2}$ of \bar{p}_1 is larger than the root $r_{2,2}$ of \bar{p}_2 and $r_{2,2} \geq r_{2,1}$. This implies that $\bar{p}_2(h) \geq \bar{p}_1(h)$ for $h \in [r_{1,1}, r_{1,2}]$. Hence for $h \in [r_{1,1}, r_{1,2}]$ (49) implies (50). Since both $\bar{p}_1(h)$ and $\bar{p}_2(h)$ are non-negative for $h \notin [r_{1,1}, r_{1,2}]$ we get the equivalence of **C6** and **C7**. This concludes the proof of Theorem 5. \diamond

Remark 3 Notice that the definition of α_i 's given in (34) implies that the left hand sides of inequalities (45)–(46) in statement **C4** are polynomials of degree 4 of variable h_{n+1} . Hence it is clear that, given h_1, \dots, h_n , the set of h_{n+1} 's leading to the variance reduction equals

the intersection of the positivity regions of $2m$ polynomials of degree 4. No transparent characterization of these sets seems available, yet for numerical values of h_1, \dots, h_n it is always possible to determine, at least numerically, which values of h_{n+1} need to be avoided to retain the VRP of the design. This opens a way to practical sequential methods of determining designs with the VRP for decreasing variances.

3.4 Proofs of Theorems 3 and 4.

Proof of Theorem 3.

By Theorem 5 it is enough to show that statement **C7** holds true.

Clearly (51) holds true for m such that $\frac{S_{1,m}}{m} < r_{1,1}$. Let m_1 be the smallest m such that $\frac{S_{1,m}}{m} \geq r_{1,1}$. Since h_i are increasing $\frac{S_{1,m}}{m}$ is increasing with m and hence we have

$$\frac{S_{1,m}}{m} \geq r_{1,1} \quad \text{for } m \in [m_1, n].$$

Inequalities (75) imply that the interval $[m_1, n]$ is non-empty, it contains at least the right-hand end point n .

Let us note that

$$\bar{p}_1 \left(\frac{S_{1,m}}{m} \right) + V_m = \left(\frac{S_{1,m}}{m} - r_{1,1} \right) \left(\left(\frac{S_{1,m}}{m} - r_{1,1} \right) + \frac{V_m}{\left(\frac{S_{1,m}}{m} - r_{1,1} \right)} - (r_{1,2} - r_{1,1}) \right).$$

From Lemmas 4 and 7 we infer that

$$\left(\frac{S_{1,m}}{m} - r_{1,1} \right) + \frac{V_m}{\left(\frac{S_{1,m}}{m} - r_{1,1} \right)} > (r_{1,2} - r_{1,1})$$

holds true for $m = n$. We will show that for $m \in [m_1, n]$ the left hand side of (3.4) is decreasing with m or, equivalently, that the expression on the right hand side of (72) is

positive. Hence, we need to show that

$$h_{m+1} < \frac{S_{2,m} - r_{1,1} S_{1,m}}{S_{1,m} - m r_{1,1}}. \quad (52)$$

Let us fix $m \geq m_1$ and define function $g(r)$ by

$$g(r) = \frac{S_{2,m} - r S_{1,m}}{S_{1,m} - m r}. \quad (53)$$

By Corollary 2 we have

$$r_{1,1}(m) = g(h_{m+1}) \leq r_{1,1} < \frac{S_{1,m}}{m}. \quad (54)$$

Notice that $h_{m+1} = g(r_{1,1}(m))$. Since for $r < \frac{S_{1,m}}{m}$ the function $g(r)$ is increasing (54) implies (52). Hence we get that

$$\bar{p}_1 \left(\frac{S_{1,m}}{m} \right) + V_m > 0$$

for any $m \in [m_1, n]$. This concludes our proof of Theorem 3. \diamond

Proof of Theorem 4.

*The design points h_i in the present case, in contrast with the case considered in Theorem 3, need be neither non-negative nor increasing and α_3 can be negative, eg for $b = -1, c = 1$. By part 1 of Theorem 5 we can rely only on the equivalence of statements **C1** – **C5**. To show that this two-point regression design has the VRP we need to consider two cases: for n even and for n odd.*

The case of even n . Condition (47) reduces in the present case to

$$p_1 \left(\frac{S_{1,m}}{m} \right) + \alpha_3 V_m = \begin{cases} \frac{1}{2} (n+1) c^2 (b-c)^2 & \text{if } m \text{ is even} \\ \frac{1}{2} \frac{m+1}{m} (n+1) c^2 (b-c)^2 & \text{if } m \text{ is odd} \end{cases}$$

$$> 0 \quad \text{for } m \in [1, n].$$

Condition (48) reduces in the present case to

$$p_2 \left(\frac{S_{1,m}}{m} \right) + \alpha_5 V_m = \begin{cases} \frac{1}{2} (n+1) (b-c)^2 & \text{if } m \text{ is even} \\ \frac{1}{2} \frac{m+1}{m} (n+1) (b-c)^2 & \text{if } m \text{ is odd} \end{cases}$$

and clearly these expressions are positive for distinct b and c and $m \in [1, n]$.

The case of odd n . Conditions (47)-(48) reduce in the present case to

$$p_1 \left(\frac{S_{1,m}}{m} \right) + \alpha_3 V_m = \begin{cases} \frac{1}{2} n \frac{(n+1)^2}{(n-1)^2} b^2 (b-c)^2 & \text{if } m \text{ is even} \\ \frac{1}{2} \frac{m-1}{m} n \frac{(n+1)^2}{(n-1)^2} b^2 (b-c)^2 & \text{if } m \text{ is odd} \end{cases}$$

$$\geq 0 \quad \text{for } m \in [1, n]$$

and similarly

$$p_2 \left(\frac{S_{1,m}}{m} \right) + \alpha_5 V_m = \begin{cases} \frac{1}{2} n \frac{(n+1)^2}{(n-1)^2} (b-c)^2 & \text{if } m \text{ is even} \\ \frac{1}{2} \frac{m-1}{m} n \frac{(n+1)^2}{(n-1)^2} (b-c)^2 & \text{if } m \text{ is odd} \end{cases}$$

$$\geq 0 \quad \text{for } m \in [1, n].$$

So, both for n even and for n odd the condition **C5** is met and hence by Part 1 of Theorem

5 we conclude that the considered two-point design has the VRP. \diamond

4 Examples and counterexamples

Example 1 Matrix W_{11} need not be positive definite, even for decreasing σ_i^2 and increasing h_i 's. Let

$$A_1 = \begin{bmatrix} 1 & 0.62 \\ 1 & 1.24 \\ 1 & 1.80 \\ 1 & 1.96 \end{bmatrix} \quad \text{and} \quad \Sigma_{11} = \text{Diag} \{1.56, 1.26, 0.78, 0.28\}.$$

Then we have

$$W_{11} = \begin{bmatrix} 0.54605 & -0.55859 \\ -0.55859 & 0.56787 \end{bmatrix}$$

and the eigenvalues of W_{11} are given by

$$-0.00174, \quad 1.11566,$$

respectively. So, the matrix W_{11} is indefinite, yet the diagonal elements are positive.

Example 2 For correlated Y_1, \dots, Y_n and uncorrelated Y_{n+1} the VRP need not hold even for decreasing σ_i^2 and increasing h_i 's. Let

$$A_1 = \begin{bmatrix} 1 & \frac{7}{10} \\ 1 & \frac{8}{5} \\ 1 & \frac{17}{10} \\ 1 & \frac{19}{10} \end{bmatrix}, \quad S = \begin{bmatrix} 2 & \frac{5}{4} & \frac{9}{10} \\ \frac{5}{3} & \frac{6}{5} & \frac{2}{5} \\ \frac{5}{4} & 1 & \frac{1}{4} \end{bmatrix}, \quad \Sigma_{11} = \begin{bmatrix} S^T S & \mathbf{0} \\ \mathbf{0}' & \frac{1}{5} \end{bmatrix},$$

where $\mathbf{0}' = [0, 0, 0]$. Then the diagonal of Σ_{11} equals $\{8.340\,278, 4.0025, 1.0325, 0.2\}$ and

$$V_n - V_{n+1} = \begin{bmatrix} -4.505\,055 & 3.277\,313 \\ 3.277\,313 & -1.987\,787 \end{bmatrix},$$

ie. adding the fourth observation results in an increase of variance for both intercept and slope estimators. We refer to Proposition 4 for a general decomposition in the case of correlated random variables.

Example 3 For uncorrelated Y_1, \dots, Y_n, Y_{n+1} the VRP need not hold even for decreasing σ_i^2 and increasing h_i 's $i = 1, \dots, n$ but with $h_{n+1} < h_n$. Let

$$A_1 = \begin{bmatrix} 1 & 0.7 \\ 1 & 1.6 \\ 1 & 1.62 \\ 1 & 1.45 \end{bmatrix} \quad \text{and} \quad \Sigma_{11} = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{4}{5} & 0 \\ 0 & 0 & 0 & \frac{1}{5} \end{bmatrix}.$$

Then we have

$$V_n - V_{n+1} = \begin{bmatrix} -0.002878 & 0.064592 \\ 0.064592 & -0.013034 \end{bmatrix}.$$

A Auxiliary results.

In derivation of our results we use the inversion formulae for block partitioned square matrices, included below for the convenience of reader:

- assuming that the inverse matrices exist (cf. Rao (1973), Problem 1.2.8),

$$\begin{bmatrix} C & B \\ B^T & D \end{bmatrix}^{-1} = \begin{bmatrix} C^{-1} + C^{-1}B(D - B^TC^{-1}B)^{-1}B^TC^{-1} & -C^{-1}B(D - B^TC^{-1}B)^{-1} \\ -(D - B^TC^{-1}B)^{-1}B^TC^{-1} & (D - B^TC^{-1}B)^{-1} \end{bmatrix} \quad (55)$$

- in the case of a nonsingular matrix C and two column-vectors B and D (cf. Rao (1973), Problem 1.2.7)

$$(C + BD^T)^{-1} = C^{-1} - \frac{1}{1 + D^TCB} (C^{-1}B) (D^TC^{-1}). \quad (56)$$

Proof of Proposition 1. *Let us note that by applying (56) and partitions (4) one gets from (9) the well known Plackett (1950) updating formula (19) for covariances of the LS estimators in the iid case, which we use in the sequel. Let us also note that*

$$WA_0^T A_0 W = \frac{aA_{00}^{-1}a^T}{1 + aA_{00}^{-1}a^T}W, \quad Wa^T aW = \frac{(aA_{00}^{-1}a^T)^2}{1 + aA_{00}^{-1}a^T}W, \quad (57)$$

$$Wa^T aA_{00}^{-1} = aA_{00}^{-1}a^T W \quad \text{and} \quad A_{00}^{-1}a^T aW = aA_{00}^{-1}a^T W. \quad (58)$$

By applying next the block-matrix inversion formula (55) and using (18) one can derive (14) with some algebra, omitted for the sake of compactness. \diamond

The following simple lemma is pivotal for the present paper.

Lemma 1 *Let $U_i, i = 1, 2, \dots, n$ be given. Inequalities*

$$\sum_{i=1}^n D_i U_i \geq 0 \quad (59)$$

hold true for all non-negative and non-increasing D_i 's if and only if

$$\sum_{i=1}^m U_i \geq 0 \quad \text{for } m = 1, \dots, n. \quad (60)$$

Proof of Lemma 1.

Clearly, by choosing $D_i = 1$ for $i = 1, \dots, m$ and $D_i = 0$ for $i > m$ we find that (59) implies (60). To show that (60) implies (59) let us note that

$$\sum_{i=1}^n D_i U_i = \sum_{j=1}^{n-1} (D_j - D_{j+1}) \sum_{s=1}^j U_s + D_n \sum_{s=1}^n U_s \quad (61)$$

and hence that the right hand side of (61) is non-negative if (60) holds true. \diamond

Remark 4 *Property (60) should not be confused with the majorization of Hardy, Littlewood & Pólya (1952) because no ordering of U_i 's is here assumed.*

Proof of Proposition 2. By (16) we have

$$\begin{aligned} W_{ii} &= \sum_{j=1}^n \left(\sum_{l=1}^k A_{00}^{-1}(i, l) A_0(j, l) \right)^2 D_j^2 - \sum_{j=1}^n \left(\sum_{l=1}^k A_{11}^{-1}(i, l) A_0(j, l) \right)^2 D_j^2 \\ &= \sum_{j=1}^n \left(\left(\sum_{l=1}^k A_{00}^{-1}(i, l) A_0(j, l) \right)^2 - \sum_{j=1}^n \left(\sum_{l=1}^k A_{11}^{-1}(i, l) A_0(j, l) \right)^2 \right) D_j^2 \geq 0 \end{aligned}$$

Hence, in order to have

$$W_{ii} \geq 0$$

for every non-increasing D_j^2 with $D_j^2 = \sigma_i^2 - \sigma_{n+1}^2$, it is sufficient and necessary, by Lemma 1, to have (22). This completes the proof. \diamond

Lemma 2 *We have*

$$\begin{aligned} \Delta_1 &= \alpha_2^2 - \alpha_1 \alpha_3 \\ &= q_{n+1}^2 (S_{2,n} - S_{1,n} h_{n+1})^2 h_{n+1}^2 \geq 0 \end{aligned} \tag{62}$$

and

$$\begin{aligned} \Delta_2 &= \alpha_4^2 - \alpha_3 \alpha_5 \\ &= q_{n+1}^2 (S_{1,n} - n h_{n+1})^2 \geq 0. \end{aligned} \tag{63}$$

Proof. The proof of Lemma 2 follows immediately from (34) and from the identity

$$(\alpha\beta - \delta\gamma)^2 - (\alpha^2 - \delta^2)(\beta^2 - \gamma^2) = (\alpha\gamma - \beta\delta)^2. \quad \diamond$$

Hence, Lemma 2 implies that both polynomials $p_1(h)$ and $p_2(h)$ have roots. We will always assume that denominators differ from zero.

Lemma 3 *Polynomials (37) and (38) have roots*

$$r_{1,1} = \frac{\alpha_2}{\alpha_3} - \frac{\sqrt{\Delta_1}}{\alpha_3} = \frac{a_1 - a_2}{a_3 - a_4} = \frac{q_{n+1}S_{2,n} - S_{2,n+1}}{q_{n+1}S_{1,n} - S_{1,n+1}}, \quad (64)$$

$$r_{1,2} = \frac{\alpha_2}{\alpha_3} + \frac{\sqrt{\Delta_1}}{\alpha_3} = \frac{a_1 + a_2}{a_3 + a_4} = \frac{q_{n+1}S_{2,n} + S_{2,n+1}}{q_{n+1}S_{1,n} + S_{1,n+1}}, \quad (65)$$

and

$$r_{2,1} = \frac{\alpha_4}{\alpha_5} - \frac{\sqrt{\Delta_2}}{\alpha_5} = \frac{a_3 - a_4}{a_5 - a_6} = \frac{S_{1,n+1} - q_{n+1}S_{1,n}}{n + 1 - nq_{n+1}}, \quad (66)$$

$$r_{2,2} = \frac{\alpha_4}{\alpha_5} + \frac{\sqrt{\Delta_2}}{\alpha_5} = \frac{a_3 + a_4}{a_5 + a_6} = \frac{q_{n+1}S_{1,n} + S_{1,n+1}}{n + 1 + nq_{n+1}}, \quad (67)$$

where Δ_1 and Δ_2 are given by (62) and (63), respectively. Moreover, we have

$$r_{1,1} = r_{2,1} = \frac{S_{2,n} - S_{1,n}h_{n+1}}{S_{1,n} - nh_{n+1}}, \quad (68)$$

$$r_{1,2} - r_{1,1} = 2nV_n \frac{h_{n+1}}{\frac{nV_n}{S_{1,n}} + \left(h_{n+1} - \frac{S_{2,n}}{S_{1,n}}\right)} \frac{(n+1)n^2V_n + n^2\left(h_{n+1} - \frac{S_{1,n}}{n}\right)^2}{n^2V_n(S_{1,n} + 2nS_{1,n} + nh_{n+1}) + n^2S_{1,n}\left(h_{n+1} - \frac{S_{1,n}}{n}\right)^2}, \quad (69)$$

and

$$r_{1,2} - r_{2,2} = \frac{2V_n \left(n \left(h_{n+1} - \frac{S_{1,n}}{n} \right)^2 + (n^2 + 1) V_n \right)}{S_{1,n} \left(\left(h_{n+1} - \frac{S_{1,n}}{n} \right)^2 + h_{n+1} \frac{V_n}{nS_{1,n}} + (2n + 1) V_n \right)} \frac{\left(h_{n+1} - \frac{S_{1,n}}{n} \right)^2 + (2n + 1) V_n}{\left(\left(h_{n+1} - \frac{S_{1,n}}{n} \right)^2 + 2(n + 1) V_n \right)}. \quad (70)$$

Let us note that Lemmas 2–6 are valid for any h_1, \dots, h_{n+1} for which the corresponding denominators differ from zero and they have been obtained by purely algebraic manipulations. For increasing non-negative sequences h_1, \dots, h_{n+1} Lemma 7 implies that the right hand sides of (68)–(69) are positive.

We shall need the following lemmas.

Lemma 4 *We have*

$$\bar{p}_1 \left(\frac{S_{1,n}}{n} \right) + V_n = nV_n \frac{(S_{2,n} - S_{1,n}h_{n+1})}{(S_{1,n} - h_{n+1}n)} \frac{\left(h_{n+1} - \frac{S_{1,n}}{n} \right)^2 + (2n+1)V_n}{S_{1,n} \left(h_{n+1} - \frac{S_{1,n}}{n} \right)^2 + (nh_{n+1} + S_{1,n}(2n+1))V_n} \quad (71)$$

Lemma 5 *For $m < n$ we have*

$$\begin{aligned} & \left(\left(\frac{S_{1,m}}{m} - r_{1,1} \right) + \frac{V_m}{\left(\frac{S_{1,m}}{m} - r_{1,1} \right)} \right) - \left(\left(\frac{S_{1,m+1}}{m+1} - r_{1,1} \right) + \frac{V_{m+1}}{\left(\frac{S_{1,m+1}}{m+1} - r_{1,1} \right)} \right) \\ &= - \frac{h_{m+1} - r_{1,1}}{(m+1) \left(\frac{S_{1,m+1}}{m+1} - r_{1,1} \right)} \left(h_{m+1} - \frac{S_{2,m} - r_{1,1}S_{1,m}}{S_{1,m} - mr_{1,1}} \right). \end{aligned} \quad (72)$$

Lemma 6 *For $m < n$ we have*

$$\begin{aligned} & \frac{S_{2,m+1} - (S_{1,m+1})h_{m+2}}{S_{1,m+1} - (m+1)h_{m+2}} - \frac{S_{2,m} - S_{1,m}h_{m+1}}{S_{1,m} - mh_{m+1}} \\ &= (h_{m+2} - h_{m+1}) \frac{m \left(\frac{S_{1,m}}{m} - h_{m+1} \right)^2 + m(m+1)V_m}{(S_{1,m} - h_{m+1}m)(S_{1,m+1} - (m+1)h_{m+2})}. \end{aligned} \quad (73)$$

We shall also need the following properties of the design points h_i which can be easily derived using convexity arguments.

Lemma 7 *If $n > 1$ and h_i are non-negative and increasing with i then $\alpha_1, \dots, \alpha_5$ are positive,*

$$h_1 < \frac{S_{1,n}}{n} < \frac{S_{2,n}}{S_{1,n}} < h_n, \quad (74)$$

$$h_1 < r_{1,1} < \frac{S_{1,n}}{n} < r_{2,2} < \frac{S_{1,n+1}}{n+1} < h_{n+1}, \quad (75)$$

and

$$\frac{S_{2,n}}{S_{1,n}} < r_{1,2} < \frac{S_{2,n+1}}{S_{1,n+1}} < h_{n+1}. \quad (76)$$

Proof of lemma 7.

Since

$$\frac{S_{2,n}}{S_{1,n}} = \sum_{i=1}^n \frac{h_i}{S_{1,n}} h_i$$

we get

$$\frac{S_{1,n}}{n} < \frac{S_{2,n}}{S_{1,n}} < \max\{h_i\} = h_n.$$

Indeed, the left hand side inequality holds because heavier weights are assigned to h_i 's of higher value. This implies (74). Moreover, we have

$$r_{1,1} = \frac{S_{1,n}h_{n+1} - S_{2,n}}{nh_{n+1} - S_{1,n}} = \sum_{i=1}^n \frac{(h_{n+1} - h_i)}{\sum_{j=1}^n (h_{n+1} - h_j)} h_i$$

and hence

$$h_1 < r_{1,1} < \frac{S_{1,n}}{n}$$

and the right hand side inequality holds because heavier weights are assigned to h_i 's of smaller values. Clearly, we have $h_1 < r_{1,1}$. Let us note that

$$r_{1,2} = \frac{q_{n+1}S_{2,n} + S_{2,n+1}}{q_{n+1}S_{1,n} + S_{1,n+1}} = \frac{q_{n+1}S_{1,n}}{q_{n+1}S_{1,n} + S_{1,n+1}} \frac{S_{2,n}}{S_{1,n}} + \frac{S_{1,n+1}}{q_{n+1}S_{1,n} + S_{1,n+1}} \frac{S_{2,n+1}}{S_{1,n+1}}.$$

Hence

$$\frac{S_{2,n}}{S_{1,n}} < r_{1,2} < \frac{S_{2,n+1}}{S_{1,n+1}}.$$

Finally, we have

$$r_{2,2} = \frac{q_{n+1}S_{1,n} + S_{1,n+1}}{n+1+nq_{n+1}} = \frac{nq_{n+1}}{n+1+nq_{n+1}} \frac{S_{1,n}}{n} + \frac{n+1}{n+1+nq_{n+1}} \frac{S_{1,n+1}}{n+1}$$

and hence

$$\frac{S_{1,n}}{n} < r_{2,2} < \frac{S_{1,n+1}}{n+1}.$$

◇

Lemmas 6 and 7 imply the following.

Corollary 2 *Let*

$$r_{1,1}(m) = \frac{S_{2,m} - S_{1,m}h_{m+1}}{S_{1,m} - mh_{m+1}} = g(h_{m+1}), \quad (77)$$

where g is given by (53). If h_i are non-negative and increasing then for $m < n$ then we have

$$0 \leq r_{1,1}(m) \leq r_{1,1}(m+1) \leq r_{1,1}(n) = r_{1,1}.$$

Corollary 3 *For h_1, \dots, h_{n+1} non-negative and increasing the values of all expressions on the right hand sides at (68) — (73) are strictly positive.*

Let us finally note that in the case of correlated random variables we have the following decomposition extending (14) onto the case of general correlated random variables. We keep here the same notation as in Proposition 1.

Proposition 4 *If the covariance matrix Σ_{11} and design matrix A_1 in a model given by (2) and (4) are of full ranks then we have*

$$V_{11} = V_{00} - \sigma_{n+1}^2 W - W_{11} + W_{22}, \quad (78)$$

where

$$W_{22} = A_{11}^{-1} (a^T \bar{\sigma}'_{1,n+1} A_0 + A_0^T \bar{\sigma}_{1,n+1} a) A_{11}^{-1}, \quad (79)$$

and $\bar{\sigma}_{1,n+1}$ is the $(n, 1)$ vector of covariances of Y_{n+1} and (Y_1, Y_2, \dots, Y_n) .

Proposition 4 can be shown in a similar way like the decomposition (14) of Proposition 1. Proposition 4 shows explicitly how the difference between V_{11} and V_{00} is affected by covariances. We also refer to Example 2, where consequences of decomposition (78) are shown and where the non-zero correlations affect the VRP.

References

- Bhaumik, D. K. & Mathew, T. (2001), ‘Optimal data augmentation for the estimation of a linear parametric function in linear models’, Sankhyā Ser. B **63**(1), 10–26.*
- Hardy, G. H., Littlewood, J. E. & Pólya, G. (1952), Inequalities, Cambridge, at the University Press. 2d ed.*
- Plackett, R. L. (1950), ‘Some theorems in least squares’, Biometrika **37**, 149–157.*
- Rao, C. R. (1973), Linear statistical inference and its applications, second edn, John Wiley & Sons, New York-London-Sydney. Wiley Series in Probability and Mathematical Statistics.*
- Sengupta, D. (1995), ‘Optimal choice of a new observation in a linear model’, Sankhyā Ser. A **57**(1), 137–153.*
- Welford, B. P. (1962), ‘Note on a method for calculating corrected sums of squares and products’, Technometrics **4**, 419–420.*